Accounting for dependence among study results in meta-analysis: methodology and applications to the valuation and use of natural resources

Florax, R.J.G.M.

2002

Faculteit der Economische Wetenschappen en Econometrie

SERIE RESEARCH MEMORANDA

Accounting for Dependence among Study Results in Meta-Analysis: Methodology and Applications to the Valuation and Use of Natural Resources

Raymond J.C.M. Florax

Research Memorandum 2002 - 5

February 2002

vrije Universiteit amsterdam

Accounting for Dependence among Study Results in Meta-Analysis: Methodology and Applications to the Valuation and Use of Natural Resources

Raymond J.G.M. Florax

Dept. of Spatial Economics, Free University,
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
Phone + 3120 4446092, Fax +31 20 4446004
E-mail rflorax@feweb.vu.nl
Website www.feweb.vu.nl/re/master-point

Abstract
Meta-analysis refers to the statistical analysis of empirical estimates obtained in previous studies, and is increasingly used in environmental and natura1 resource economics as a complement to a state-of-the-art literature review. The occurrence of dependence or auto-correlation among study results, for multiple estimates from the same study or for estimates from different studies, is a compelling problem that is usually ignored. This paper suggests that autocorrelation tests and estimators developed for other types of data constitute an appropriate solution to measuring and remedying dependence in meta-analysis. Moreover, visualization by means of a scatterplot provides a useful tool for the interpretation of dependence, and helps to detect outliers. The paper provides illustrations of the techniques through meta-analyses on the valuation of wetlands and the price elasticity of residential water demand. The applications show that between-study dependence is usually sufficiently modeled by means of variability in study characteristics. Ignoring within-study dependence, however, can result in biased estimators and makes inferences from meta-analyses imprecise in size and significance.

JEL: C12, C13, D12, Q25
Keywords: meta-analysis, autocorrelation, dependence, heterogeneity

1. Introduction

Meta-analysis is by now a well-accepted tool in environmental and natura1 resource economics. It complements the conventional state-of-the-art review of the literature by providing a statistical analysis of empirical results obtained in previous studies. Nelson (1980) was the first to use the technique, assessing the average Noise Depreciation Index over studies, in addition to presenting a qualitative survey of property value studies estimating the impact of airport noise. Subsequently, between 1980 and 2001 approximately 40 meta-analyses appeared in environmental and natura1 resource economics, half of them addressing the valuation of pollution and recreation, and one-third being concemed with the nexus of agriculture, land use, and the use of natura1 resources (Florax 2002b). Some of the more prominent valuation studies are Katzman (1987), Smith and Huang (1993, 1995), Schwartz (1994), Loomis and White (1996), Espey and Kaufman (2000), and Woodward and Wui (2001). The demand for derivates of natura1 resources, such as gasoline and water, is covered in Espey (1996, 1998), Espey et al. (1997), and Dalhuisen et al. (2001), among others. Methodologically oriented issues, such as differences arising from the use of hypothetical or revealed preferences, are addressed in, for instance, Carson et al. (1996), and List and Gallet (2001).

Meta-analysis was developed in the experimental context of agronomy and medicine. The traditional experimental set-up features two large sample groups, one of which receives treatment (the "experimental" group) and the other does not (the "control" group). The treatment effect can then be straightforwardly isolated as a standardized mean difference between groups. An advantage of this approach is that the experimental set-up is rather homogeneous over different experiments, and effect size indicators do not depend on the unit of measurement.[1] An extensive literature on meta-analysis techniques in an experimental setting has been developed (see, Hedges and Olkin 1985; Cooper and Hedges 1994). In the largely non-experimental set-up in economics, the effect size indicator is typically an elasticity or a nomina1 value, such as consumer surplus or willingness to pay. Elasticities are often derived as point

[1] The effect size can be measured as a standardized mean difference, a correlation coefficient, a difference in proportions, an index of proportion of variance accounted for, or a similar statistical summary indicator. Most effect size indicators can be derived from each other. There is a highly specialized, voluminous literature on this topic (see, for instance, Rosenthal 1984).

estimates of log-linear demand models, estimated by means of econometric techniques. The advantage of using point elasticities is that the effect size does not depend on the unit of measurement, and its distribution is known to be asymptotically normal.

Meta-analysis is plagued by three methodological problems (see also Glass et al. 1981; Stanley 2001): selection and publication bias, heterogeneity among studies, and dependence of study results. Selection effects occur when the process of literature retrieval is such that the likelihood of sampling a study is correlated with the effect size measure. This may be due to restrictive sampling over time, within a country or language zone, or alternatively because of a focus on a specific theoretical or modeling approach. A special case of selection effects is caused by researchers self-censoring the publication of 'negative' or statistically insignificant effects, a practice that may be invigorated by editorial selection processes. Combining and explaining published effects that constitute a biased sample of the population's "true" effect, is detrimental to the validity of meta-analysis as a summarizing technique (Card and Krueger 1995; Ashenfelter et al. 1999).

Heterogeneity among studies has many dimensions. Studies may differ according to, for instance, quality of the research design and data, type of data, estimator, functional form and specification of the model, and underlying theory. These differences can be included in the meta-model, either as fixed observable effects or as a random unobservable variate. In addition to this substantive heterogeneity, the distribution of effect sizes is inherently heteroscedastic, because estimated effect sizes are based on studies with different sample sizes. Heterogeneity is in most meta-analyses treated adequately by specifying a fixed or random effects model (see Schwartz 1994; Jeppesen et al. 2001), and the application of either a weighted regression approach (Cavlovic et al. 2000) or a heteroskedasticity robust variance estimator (Woodward and Wui 2001).[2]

The problem of lacking independence has not been addressed sufficiently in economic meta-analyses. Few studies refer to the potentially disturbing influence of correlated effect sizes, although the occurrence of dependence is much more likely in

[2] An interesting alternative is the use of hierarchical multilevel modeling (Brouwer et al. 1997).

economics – as compared to, for instance, medicine – because of multiple sampling of estimates per study. Statistical testing for autocorrelation is non-existent. Espey (1996) and Espey and Thilmany (2000) hint at the use of a dependence test by computing cross-correlation coefficients among residuals of the same study to attain an indication of within-study dependence.[3] Smith and Kaoru (1990a,b), Smith and Huang (1993, 1995), Boyle et al. (1994), and Smith and Osborne (1996) do not test for dependence, but they use estimators that allow for cross-correlation within studies.[4]

In this paper we present an approach to test for correlation within and between studies in meta-analysis, and subsequently estimate meta-regression models taking into account the dependence when it occurs. Section 2 presents the typical set-up for meta-analysis in (environmental) economics, and concisely covers pivotal assumptions concerning heterogeneity and dependence. In Section 3, the specification and interpretation of within- and between-study dependence is discussed. Section 4 introduces statistical tests and estimators for autocorrelated data in a meta-analysis context. In Section 5, the tests for dependence within and between study results are linked to a visual inspection tool that is of considerable practica1 relevance for the meta-analyst. In Section 6, the use of tests, visualization tools and estimators is illustrated by reanalyzing two recent applications. One is concemed with the valuation of wetlands and the other with price elasticities of residential water demand. Section 7 of the paper provides conclusions.

2. The set-up of meta-analysis in (environmental) economics

The starting-point for meta-analysis in economics is usually a series of observations T_{ij} on the population effect size θ_{ij}, with associated standard error σ_{ij}, for estimates j (= 1, 2, ... J_i) sampled from studies i (= 1, 2, . . . , I). This set-up shows that multiple

[3] Jeppesen et al. (2001) mention that there is no dependence among the residuals, but it remains unclear how this statement is substantiated.

[4] Smith and Osborne (1996) apply weights defined by the number of sampled estimates from a study in order to give each study equal weight and mitigate the influence of dependence. This procedure, however, only reduces potential problems with heteroskedasticity, but does not affect the impact of dependence. There is also some confusion as to how the cross-correlation is implemented, as Greene (1993, p. 453) seems to imply that the data should form a balanced panel in order to avoid computational problems.

sampling from the same study, which is common practice in (environmental) economics, results in a sample with pooled data. It is important to note that the data do not, however, form a panel. The ordering of estimates within studies is arbitrary, so j is an indicator without substantive meaning. Sampling can also be such that single estimates of studies (i.e., $J_i = 1$) are combined with multiple estimates from other studies (i.e., $J_i > 1$).

Several models can be distinguished depending on the heterogeneity that is allowed for in the meta-analysis. The simplest model assumes that the underlying population effect size is the same for all studies and estimates, $\theta_{ij} = \theta \ \forall i, j$. Allowing the population effect size to differ among studies (maintaining homogeneity within each study) introduces somewhat more heterogeneity, $\theta_{1\bullet} \neq \theta_{2\bullet} \neq \cdots \neq \theta_{I\bullet}$. In addition it is customary to hypothesize that part of the variation among effect sizes can be attributed to various identifiable study characteristics. These characteristics can be modeled by means of dummy variables ('fixed effects'), but also as interval (for instance, a time trend) or ratio scale variables (for instance, GDP per capita). Somewhat confusingly all the above models are referred to as *fixed effects* models in the meta-analysis literature.

There has been considerable debate on whether it is appropriate to assume that the heterogeneity can be fully explained by means of 'fixed effects' (see Sutton et al. 2000, pp. 83-84). It is often argued that it is preferable to assume that the underlying population effect sizes differ between studies, and that the studies' population effect sizes can be seen as random draws from a normal distribution. In the meta-analysis literature this model is referred to as the *random effects* model. If, in addition, some of the variation is modeled through additional exogenous variables (as above), the meta-analysis literature uses the term mixed *effects* model.

In environmental economics virtually all meta-analyses are based on the fixed effects modeling approach. Exceptions are Schwartz (1994), who employs a random effects model, and Brouwer et al. (1997) and Jeppesen et al. (2001), who use a mixed effects model among other models.

A typical assumption of all meta-analyses in environmental economics, usually left implicit, is the assumption that $\mathrm{Cov}[T_{ij}, T_{kl}] = 0$, for $i \neq k$ or $j \neq l$. Hence, despite multiple sampling and pooling of data, the effect size estimates are taken to be independently distributed or *not* autocorrelated.[5] This is likely to be an unwarranted assumption because within a study the estimates are derived (largely) by employing the same data and similar models, and between studies there are overlapping similarities in, for instance, space-time coverage, research design, type of data, and specification and estimation procedures, that may not be accounted for through exogenous variables.

The structure of the autocorrelation or dependence in meta-analysis is not identical to what we know from the time series domain, where an a priori structure exists. Autocorrelation in time series implies a unidirectional causation pattem going from past to present. Familiar autocorrelation tests, such as the Durbin-Watson test, and time series estimators for autoregressive and moving average models (or more complicated forms) are therefore inappropriate for meta-analysis. Even a cautious use of the familiar Durbin-Watson test should be avoided, because inferences are likely to be misleading owing to the one-sided comparison.

3. Autocorrelation in meta-analysis

Autocorrelation in meta-analysis is much more akin to network correlation in social networks, spatial correlation among regions or countries, or clustered diffusion pattems of contagious diseases as studied in epidemiology. In those cases the autocorrelation is multidimensional. Multidimensionality implies that an observation can be influenced by multiple other observations in the sample (which in time series is equivalent to a distributed lag), but also that the influence is two-sided. Instead of the unidimensional lag operator L in time series that shifts back an observation y over k periods in time, $y_{t-k} = L^k y_t$, a multidimensional lag operator is needed in the context of meta-analysis. Following an analogous concept developed in spatial statistics, such a multidimensional lag operator can be formalized as $L^{kl}T_{ij} = \sum_k \sum_l w^S_{ij,kl}T_{kl}, \forall k, l \in S$, where L^{kl} is the lag operator associated with similarity class S that identifies the set of effect sizes potentially linked to estimatej of study *i*. The elements $w^S_{ij,kl}$ specify a set of weights, for instance through a binary zero-one indicator (Cliff and Ord 1981;

[5] The terms ‘autocorrelation’ and ‘dependence’ are used interchangeably in this paper, although auto-correlation is somewhat stricter in that it also presupposes that the variable considered is normally distributed.

Cressie 1991), so that potentially dependent estimates l of studies k are compared to estimate j of study i.

Potentially, all observations can belong to one and the same similarity class. As a result, however, everything depends on everything else, and the system is not identifiable unless an exogenously provided decay pattern (such as by distance, in the spatial case) can be provided. Such a 'natura1 structure' does, however, not exist for meta-analysis, and as a consequence, exogenous structure has to be imposed. A new approach to impose such an exogenous structure is to distinguish "within" from "between" study effects. Within-study effects can be defined by means of the similarity class that includes all observations sampled from the same study. The definition of this class is in most cases exogenous. We can then define weights for within-study autocorrelation as:

$$w_{ij,kl}^{W} = \begin{cases} 0 & i = k, j = l \\ \dfrac{1}{J_i - 1} & i = k, j \neq l \\ 0 & i \neq k \end{cases} \quad . \tag{1}$$

These weights are zero except when two different estimates, j and l, are sampled from the same study $(i = k)$. The specification of non-zero weights is such that an effect size from a specific study is compared to the average of the other effect sizes from the same study.

Between-study effects can be determined using diverse similarity classes, for instance referring to similarity in theoretical or modeling perspective, type of data, type of estimator, or space-time coverage. The main problem with such an approach is that the weights are not exogenous. We therefore suggest a specification of weights defined in terms of the number of sampled studies I, and the number of estimates sampled from each study, J_i . In effect this secures that the weights are exogenously defined. We define the weights for between-study effects as:

$$w_{ij,kl}^{B} = \begin{cases} 0 & i = k \\ \dfrac{1}{J_i (I - 1)} & i \neq k, j \neq l \end{cases} \quad . \tag{2}$$

The between-study weights are different from zero except when two estimates, j and l, are sampled from the same study $(i = k)$. The specification of the non-zero weights is such that an effect size from a specific study is compared to the weighted average of the (estimated) means of the other studies.[6]

Because the notation is rather cumbersome, owing to the pooled nature of the data, we provide a simple numerical example. Consider a series of six effect sizes $T_{ij} = \{0.1, 0.3, 0.2, 0.5, 0.3, 0.15\}$ taken from three studies. Three estimates are taken from study 1, two estimates from study 2, and one estimate from study 3. Matrices representing the within- and between-study structure are straightforward to derive using the above definitions, as:

$$
(3) \qquad \mathbf{W}^{W} = \left[\begin{array}{ccc|cc|c} 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \text{ and } \mathbf{W}^{B} = \left[\begin{array}{ccc|cc|c} 0 & 0 & 0 & 0.25 & 0.25 & 0.5 \\ 0 & 0 & 0 & 0.25 & 0.25 & 0.5 \\ 0 & 0 & 0 & 0.25 & 0.25 & 0.5 \\ \hline 0.167 & 0.167 & 0.167 & 0 & 0 & 0.5 \\ 0.167 & 0.167 & 0.167 & 0 & 0 & 0.5 \\ \hline 0.167 & 0.167 & 0.167 & 0.25 & 0.25 & 0 \end{array}\right]
$$

The matrices are standardized by definition, implying that the row sums equal one, except for zero-rows in the within-study weight matrix. The latter refer to studies with one sampled estimate, for which there is obviously no within-study correlation.

By means of the above matrices we construct $(n \times 1)$ vectors of the relevant within- and between-study averages, based on the original (nxl) column vector $\mathbf{t}$ comprising the effect sizes T_{ij} (n is the total number of observations). We define:

$$
(4) \qquad \mathbf{t}^{W} = \mathbf{W}^{W}\mathbf{t} = \left[\begin{array}{c} 0.25 \\ 0.15 \\ 0.20 \\ \hline 0.30 \\ 0.50 \\ \hline 0 \end{array}\right] \text{ and } \mathbf{t}^{B} = \mathbf{W}^{B}\mathbf{t} = \left[\begin{array}{c} 0.275 \\ 0.275 \\ 0.275 \\ \hline 0.175 \\ 0.175 \\ \hline 0.300 \end{array}\right] \text{ where } \mathbf{t} = \left[\begin{array}{c} 0.10 \\ 0.30 \\ 0.20 \\ \hline 0.50 \\ 0.30 \\ \hline 0.15 \end{array}\right]
$$

[6] Note that this provides an approximation to between-study correlation, because *each* effect size estimate of a particular study is compared to the weighted average of the means of the effect sizes of the other studies. This is strictly speaking not exactly identical to comparing mean effect sizes between studies. In Section 5 we return to this issue.

It is apparent from comparison of the vectors that each element of $\mathbf{t}$ has the average of the other estimates within the same study as corresponding elements in $\mathbf{t}^{W}$, and the weighted average of the means of the other studies in $\mathbf{t}^{B}$. The last element of $\mathbf{t}^{W}$ is zero, showing again the absence of within-study correlation for studies with one sampled estimate only.

4. Testing for autocorrelation and estimation of autoregressive models

In his seminal work on statistical maps Moran (1948, 1950) suggests to measure the degree of dependence, for interval and ratio scale data, through the statistic I. In matrix notation and adapted to the meta-analysis context, the statistic reads as:

(5) $$I = \frac{n}{S_0} \cdot \frac{\bar{\mathbf{t}}' \mathbf{W}^{S} \bar{\mathbf{t}}}{\bar{\mathbf{t}}' \bar{\mathbf{t}}},$$

where $\bar{\mathbf{t}}$ is the vector of observed effect sizes measured in deviations from the grand sample mean of T_{ij}, $\mathbf{W}^{S}$ the weight matrix applying to a specific similarity class, and S_0 the sum of the elements of the weight matrix. For a between-study weight matrix the scaling factor n/S_0 equals one. For a within-study weight matrix n/S_0 scales the statistic for the 'missing' covariances of single estimate studies included in the meta-sample.

Moran's I is a special case of the genera1 cross-product statistic derived by Hubert et al. (198 1; see also Getis 1991 for an overview). The statistic compares the covariance among an exogenously defined set of effect sizes to the variance of all observed effect sizes. It can conveniently be interpreted as a correlation coefficient, although it is not necessarily bounded to the [-1,+1] interval, and it is centered about $-1/(n-1)$ instead of zero. Values greater than the theoretical mean signal the occurrence of similar effect sizes within or between studies (either high or low values). Values smaller than the mean indicate the joint occurrence of high and low effect sizes within or between studies. A value not significantly different from the mean can be taken as evidence of a random distribution of effect sizes within or between studies. In that case, the value of a specific estimated effect size could have been observed for any study i and estimate j.

It can be shown that Moran's I is normally distributed, $I \sim N(\mu,\sigma)$, so a forma1 test of the null hypothesis that the effect sizes are independently distributed can be based on z,, the standardized value of I, which follows a standard normal distribution. Cliff and Ord (1981) theoretically derived the moments of Moran's I under the assumption that T is normally distributed. If the distribution of T is unknown or does not correspond to the normal, the distribution can be approximated in a nonparametric framework using a randomization approach or empirically generated using a permutation approach (Cliff and Ord 1981, pp. 42-46 and 63-65). Simulation experiments by Cliff and Ord (1981) show the test performs reasonably well, even in smal1 samples.

Moran's I for regression residuals is identical to the formulation in (5), replacing the vector of effect sizes with the vector of OLS residuals.[7] A disadvantage of Moran's test for regression residuals is the very genera1 alternative hypothesis simply specifying correlated residuals due to any cause. Lagrange Multiplier (LM) tests, developed in a maximum likelihood framework, are more attractive because they are explicitly linked to specific alternative models. In the spatial econometric literature, two major altemative models are distinguished (Anselin and Bera 1998; see Anselin 200 1, for more complicated models).

One model, referred to as the error ***model,*** recognizes that (erroneously) omitted variables can be autocorrelated. This model does not have a substantive interpretation and reads, in the context of meta-analysis, as:

$$\mathbf{t} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ and } \boldsymbol{\varepsilon} = \lambda \mathbf{W}^{S}\boldsymbol{\varepsilon} + \boldsymbol{\mu}, \tag{6}$$

where λ is the autoregressive parameter indicating the magnitude of the unspecified dependence within or between studies, and $\boldsymbol{\mu}$ is a well-behaved error term. The appropriate LM test is identical to a scaled Moran coefficient (Burridge 1980), and takes on the form:

[7] Because least squares residuals are correlated by definition, the moments of Moran's 1 implied by (5) are not appropriate for a test whether the error term is correlated. The appropriate moments in a regression framework, assuming a normal independent distribution for the errors, are derived in Cliff and Ord (1972).

(7) $$LM_\lambda = \frac{1}{c}\left(\frac{\mathbf{e}'\mathbf{W}^S\mathbf{e}}{s^2}\right)^2$$

where s^2 is the maximum likelihood variance $\mathbf{e}'\mathbf{e}/n$, and $c = \mathrm{tr}(\mathbf{W}^{S'}\mathbf{W}^S + \mathbf{W}^{S^2})$, with tr as the matrix trace operator. The test asymptotically follows a χ^2 distribution with one degree of freedom. Anselin (1988a) shows that ignoring the correlated error structure does not result in biased estimates, but the estimates are inefficient.

The other model, referred to as the *Zag* model, includes a lagged dependent variable among the regressors, because the observations on the dependent variable can be realized simultaneously. The model reads as:

(8) $$\mathbf{t} = \rho\mathbf{W}^S\mathbf{t} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu},$$

where ρ is the autoregressive parameter indicating the magnitude of the dependence of effect sizes within or between studies. The appropriate LM test has the same asymptotic distribution as the error test, and looks similar (Anselin 1988b):

(9) $$LM_\rho = \frac{1}{nJ_{\rho\cdot\beta}}\left(\frac{\mathbf{e}'\mathbf{W}^S\mathbf{t}}{s^2}\right)^2,$$

where $J_{\rho.\beta} = [(\mathbf{W}^S\mathbf{X}\mathbf{b})'\mathbf{M}(\mathbf{W}^S\mathbf{X}\mathbf{b}) + cs^2]/ns^2$ is part of the estimated information matrix, M the projection matrix $(1 - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')$, and b the OLS parameter vector. Ignoring the endogeneity issue implied by the lag model is more serious than ignoring a correlated error term, because the OLS estimator is biased as well as inconsistent.

The lag model in (8) can be rewritten as $\mathbf{t} = (1-\rho\mathbf{W}^S)^{-1}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu})$, showing that both the error and the lag model have a correlated error structure. This explains why Moran's I has substantial power against both alternatives, and is therefore not very useful as a misspecification test for regression models. Anselin and Florax (1995) and Anselin et al. (1996) show that the LM tests have considerable power to detect the right model specification, even in smal1 samples. Appropriate maximum likelihood estimators for the error and the lag model, and extensions incorporating (groupwise) heteroskedasticity are given in Anselin (1988a, 1992).

5. Visualization of dependence

In meta-analyses outside the realm of economics visualization tools are frequently used to investigate the heterogeneity of effect size estimates, in addition to statistical testing. The so-called Q test, where Q is defined as a (conditional) variance weighted deviation of the average effect size, is graphically depicted in a Galbraith diagram (Cochran 1954; Galbraith 1988). The Q statistic is 'global' or 'overall' in the sense that it applies to all observations. Hedges and Olkin (1985) show that each study's (or observation's) contribution to the overall statistic can be formalized by means of a 'local' statistic q. The Q test should be used cautiously, among other things because its power is low (Sutton 2000). Further discussion of the test is beyond the scope of this paper, but it is interesting to note that in spatial statistics and geostatistics a 'local' version of Moran's I with a concurrent scatterplot have been developed (Anselin 1995; Cressie 1991). In this section we will demonstrate how the local statistic and the scatterplot can be fruitfully applied in economic meta-analyses.

Following Anselin (1995) the local Moran for an individual estimate j of study i can be expressed as:

(10)
$$I_{ij} = \frac{n}{S_0^{ij}} \cdot \frac{\bar{T}_{ij}\mathbf{w}_{ij}\bar{\mathbf{t}}}{\bar{\mathbf{t}}'\bar{\mathbf{t}}},$$

where the bar indicates that variables are measured in deviations from the overall sample mean, $\mathbf{w}_{ij}$ is the row of the weight matrix pertaining to estimate j of study i, and S_0^{ij} refers to the sum of the weights in $\mathbf{w}_{ij}$. Anselin (1995, p. 99) derives the moments of the local Moran under the null hypothesis of independence, but points out that statistical inference is safest when taking a randomization approach, because the exact distribution of the statistic is stil1 unknown. The global Moran coefficient is equal to the sum of the local Moran coefficients, up to a proportionality factor defined in terms of S_0 (Anselin 1995 provides details).

Similar to the use of the local q statistic, the local Moran can be used to identify influential estimates and clusters of similar values. This can also be achieved through the use of a scatterplot. As with any statistic expressed as a ratio of a quadratic form and its sum of squares, Moran's I is equivalent to a bivariate regression coefficient of a regression of **Wt** on **t**, and can be visualized in a scatterplot (Anselin 1996).

The Moran scatterplot has standardized effect size values on the horizontal axes, and standardized values of **Wt** on the vertical axes. The slope of a linear trend line in a plot of all observations corresponds to the global Moran coefficient, given the abovementioned equivalence. Because of the standardization one can easily judge clustering of (dis)similar values from the scatterplot. The upper-right and lower-left quadrants show observations with above and below average values that contribute positively to the overall autocorrelation because their local Moran is positive. The upper-left and lower-right quadrants show dissimilar values that contribute negatively to the overall autocorrelation because their local Moran is negative. Finally, one can use the scatterplot to identify outliers, for instance, those observations that are further than two standard deviations away from the overall sample mean (represented by the origin) .

We point out some significant details of the Moran scatterplot for the numerical example given in Section 3. Figure la provides the scatterplots of within-study correlation for the numerical example. The top graph shows the estimated effect sizes connected by a trend line for estimates from the same study. Study 3 is a single estimate study. Given the definition of weights (in effect evoking a comparison of averages) the trend lines per study are typically downward sloped. The bottom graph shows the estimates of all studies, with a trend line added. It is important to note that for within-correlation the slope of the trend line is not necessarily equal to the global Moran's I. In this example the Moran coefficient for within-dependence is 0.24 and the slope of the trend line is 0.20. This deviation is caused by the inclusion of single study estimates. Figure la also shows that overall there is positive autocorrelation within studies. Relatively high effect sizes as well as relatively low effect sizes are clustered within studies (in this case in study 2 and 1, respectively).

Figure lb presents scatterplots for between-study dependence. The graphs show that between studies the effect sizes are dissimilar, given the concentration of points in the upper-left and lower-right quadrant. The overall dependence is -0.30, which corresponds exactly to the slope of the trend line in the middle graph. For between-study correlation the correspondence of the slope of the trend line to Moran's

I holds, because the weights matrix does by definition not contain rows with only zeros and all rows sum to one.

The third graph of Figure lb (bottom) presents study means only in order to demonstrate that the procedure we suggest to measure between-study dependence is an approximation (see footnote 6). The exact measurement, using only study means, is -0.5 whereas 'our' Moran coefficient is -0.3. The difference is due to the assignment of the weighted average of the means of the other studies to *each* estimate of a specific study. This is easily verified in the top graph where the between-averages for each estimate of a specific study are located on a horizontal line. Three arguments are of paramount importance for the justification of the approximate procedure. First, it is not possible to judge the significance when using study means only, unless one assumes that the study means are independent (in which case the usual t-statistic applies). Second, accurate modeling of between-study dependence is only feasible in a hierarchical modeling set-up where potentially the error terms for estimates as well as for studies are autocorrelated. This constitutes a fairly complicated hierarchical model, for which to date no estimators are available. Finally, the accuracy of the variant we suggest depends on the proportion of single study estimates in the sample, and is likely to have a reasonable asymptotic accuracy (with increasing sample size).

6. Applications

We demonstrate the relevance of taking into account dependence within and between effect size estimates of a series of studies by re-analyzing two recent meta-analyses. The applications reflect two different types of analysis typical for environmental and natura1 resource economics. One study uses values in constant prices as effect size estimator, and the other uses price elasticities.

The meta-analysis by Woodward and Wui (2001) analyzes per acre values of wetland (in constant 1990 US dollars). The effect size is measured as a per acre value in constant prices, but due to the research design of (some of the) original studies, no information is available on the estimated standard errors of the effect sizes.[8] The

[8] The authors provide the complete data set, including a description of each study and an explanation of the interpretations of the data, online at http://ageco.tamu.edu/faculty/woodward/.

other study is a meta-analysis of price and income elasticities of residential water demand by Dalhuisen et al. (2001). For the subset of point-elasticities estimated by means of double-log specifications the estimated standard errors of the effect sizes are known. We use a subset of the price elasticities of the dataset that have been selected using the criterion of availability of information about the tariff structure, in particular decreasing, flat or increasing block rate pricing.[9]

In the wetlands meta-analysis 65 observations are available and in the water demand analysis 123. Figure 2 presents the distribution of effect sizes and the sampling distribution according to studies. For the wetlands analysis we follow Woodward and Wui's choice of using the natura1 logarithm of the estimated effect size as the variable of interest. The assumption of a log-linear relation is not uncommon in economic meta-analyses, and reflects nonlinearity in addition to contributing to remedying potential heteroscedasticity and non-normality. The mean of the natura1 log of the per acre value of wetland is 4.95 (corresponding to approximately 140 US$ of 1990), with a standard deviation of 2.28. For the water demand application Figure 2 shows strong evidence of water demand being inelastic and negative in response to price changes. The mean price elasticity is -0.52, with a standard deviation of 0.49. A slight indication for heteroskedasticity is associated with the increasing standard errors for relatively large effect sizes.[10] This may also be indicative of publication bias, because estimates with relatively large standard errors will only be significant if the effect size is large in magnitude (see Florax 2002a, for details). The bottom of Figure 2 shows that the effect sizes for wetlands are sampled from 33 studies, and those for water demand from 27 studies, with the wetland sample having substantially more single study estimates and on average a lower number of sampled estimates per study (two as compared to almost five for the water application).

---

[9] The complete, annotated database for the residential water demand study is available at http://www.feweb.vu.nl/re/master-point. Modified databases for both applications presented in this paper, including weight matrices, numerical examples and estimation output, are available from the same location.

[10] Estimated standard errors are available for approximately one-third of the effect sizes. The graph for price elasticities in Figure 2 contains one elasticity with an associated standard error that is extreme and goes beyond the boundaries of the graph. The elasticity is -1.2 with a standard error of 5.11.

The hypothesis that the effect size variable is normally distributed cannot be rejected for the wetlands application as opposed to the water demand case. The Wald tests are 1.689 and 856.355, with corresponding p-values of 0.43 and 0.00. In testing for overall dependence we therefore apply the asymptotic normal distribution assumption for the wetlands case and the randomization assumption for the water demand application.

We assess both within- and between-study dependence. There is strong evidence for significant positive within-study dependence. Moran's I for the wetland data is 0.49 (z = 3.50), and for the water use data 0.28 (z = 4.53), indicating that within studies similar values, either high or low, are found together. The test results for between-study dependence indicate that significant negative autocorrelation is present, although it is rather smal1 in magnitude. For the wetland application we find -0.02 (z = -3.25) and for the water application -0.02 (z = –2.70), for the Moran I test.[11] The direction and significance of these results are likely indicative of most meta-analyses in economics, because the characteristics of the research design have only limited sampling variation – if any (see also Hedges 1997). We also assess the accuracy of our approximation of between-study dependence by calculating the correlation between studies strictly for study means (see Section 5). In the wetlands and water demand cases we find -0.0242 and -0.0160 for our approximation versus -0.0268 and -0.01492 for the exact approach, respectively. So we can conclude that the approximation is sufficiently accurate in medium-sized samples. It is obvious that the higher the proportion of single study estimates the better the approximation will be. Hence, for smal1 meta-sample sizes with a high proportion of single study estimates the accuracy of the approximation is probably sufficient as well.

Figure 3 presents Moran scatterplots for both applications and within- and between-study dependence. The graphs for the wetland application show that all observations are within two standard deviations of the sample mean and that the width of the 'cloud' does not change along the trend-line. Hence, there are no outliers and the variance is constant (homoskedasticity).

---

[11] Similar results are found using empirically generated distribution functions based on 10,000 permutations, although the signifícance levels are slightly lower in most cases.

The slopes of the trend-lines for both applications correspond exactly to Moran's I for the between-study dependence. For within-study dependence they are not accurate, due to single study estimate sampling. One should note that for the water example the slope of the trend-line is much closer to the Moran coefficient (I = 0.278, slope = 0.269) than for the wetland case (I = 0.49, slope = 0.33), owing to the much higher proportion of single study estimates in the wetland example (0.64 versus 0.15).

The graphs for the water demand application show the same clustering pattern of high-low/low-high values, with a corresponding negative Moran coefficient for between-study dependence, and a low-low/high-high pattem, with a corresponding positive Moran, for within-study dependence. In the water case, however, the graphs show clear signs of negative outliers and heteroskedasticity (non-constant variance along the trend-line). [12] There has been much debate in the meta-analysis literature about the proper treatment of extreme observations (of either sign or size). Although some authors maintain that extreme heterogeneity precludes combining study results because it amounts to 'combining apples and oranges and the occasional lemon' (Furberg and Morgan cited in Sutton 2000, p. 53), others maintain that removing outliers and extreme results at an early stage of the meta-analysis should be avoided because it can introduce (substantial) bias into the meta-results. The influence of removing extreme results should instead be explored in a sensitivity analysis (Sutton 2000; Stanley 2001).

Woodward and Wui (2001) regress the natura1 log of the value per acre wetland converted to 1990 US dollars on the year of the study (Year, 1960 = 0), the natura1 log of the size of the wetland (ln Acres), whether the wetland is a coastal wetland (Coastal), the services of the wetland, and study characteristics. Wetland services are defined as: reduced damage due to flooding and severe storms (Flood), increased water quantity (Quantity), reduced costs of water purification (Quality), improvements in downstream recreational (Rec. Fish) and/or commercial fisheries (Com. Fish), hunting (Birdhunt) and observation (Birdwatch) of wildlife, amenity

[12] Compared to the initially available samples Dalhuisen et al. (2001) excluded one outlier and seventeen positive price elasticities a priori, and Woodward and Wui (2001) excluded one outlier, one other value estimated using energy analysis, and five values estimated using the market value of the output.

value provided by proximity to the environment (Amenity), nonuse appreciation of species (Habitat), and erosion reduction (Storm). Study characteristics include whether the results are published (Publish), whether it is an estimate of producer surplus (PS), and whether it is estimated by means of hedonic pricing (HP), net factor income (NFI), replacement cost (RC), or the travel cost (TC) method. The quality of the primary studies is incorporated through (subjective) identification of studies of questionable quality in terms of data (Data), theory (Theory), and econometrics (Metric).

Table 1 presents the results for three different specifications presented in Woodward and Wui (2001). The first three columns are an exact replication of their results, except that standard errors are not White-adjusted because heteroskedasticity is not present according to the Breusch-Pagan test. The results of the misspecification tests indicate that the extended model provides a reasonably good fit, and there are no apparent difficulties with multicollinearity, normality or the error term, and heteroskedasticity.

The between-studies dependence tests reveal that heterogeneity between studies is adequately modeled, so that there is no apparent dependence that is left unmodeled. Within studies the estimates are, however, autocorrelated. The positive sign of Moran's I demonstrates that within studies high values, or alternatively low values, are clustered. The LM tests clearly point in the direction of a lag model as the appropriate altemative.

The second set of three columns in Table 1 gives the results of the lag model, with the autoregressive term being significantly different from zero. The Breusch--Pagan test points to homoskedasticity as the correct alternative, and the Likelihoods indicate that the lag models taking into account autocorrelation within studies are preferable to the initial OLS results. Most important is the considerable bias of the estimates of the initial specifications. It should be noted, however, that the coefficients of the lag model do not represent marginal effects, because instead of $\partial \mathbf{y} / \partial \mathbf{x} = \beta$ for a model without lagged variables, $\partial \mathbf{y} / \partial \mathbf{x} = [(1 - \rho \mathbf{W})^{-1}]' \beta$ for the lag model. The latter is equivalent to $[(1 - \rho \mathbf{W})']^{-1} \beta$, which amounts to β multiplied by the column sums of the transformation matrix $(1 - \rho \mathbf{W})^{-1}$. As a result, the marginal effect is no longer uniform between studies. Woodward and Wui's estimate of a 2.9%

fall in value (see their Model C) for a ten-acre wetland due to a 1% increase in size, varies between 1.8% for studies with one estimate and 2.6% for studies with multiple estimates, and is on average 2.1% for the extended lag model.

Table 2 presents the results for one of the Dalhuisen et al. (2001) meta-models for price elasticities. Price elasticities of residential water demand are regressed on characteristics of the theoretical and modeling approach of the original studies (use of average/fixed prices, Shin prices, conditioning on income, inclusion of a difference variable, and simultaneous modeling of discrete and continuous choices) as well as notable features of the elasticity (long-run, segment, and increasing or decreasing block rate pricing). In addition the analysis catches variation over space and time (Western US and time-trend), and is conditioned on the per capita GDP leve1 as well as on characteristics of the specification (accounting for household size and season) and the data (monthly data, panel data, and summer data) of the original studies.

Dalhuisen et al. (2001) report the White-adjusted OLS results. The condition number (CN) for multicollinearity is relatively high, and the Jarque-Bera test (JB) on normality of the errors is very significant. We do not treat these two aspects any further here, but instead concentrate on heteroskedasticity and dependence. Heteroskedasticity is clearly present, and for comparative purposes we therefore provide the results of the weighted least squares (WLS) estimator. For the WLS estimator we take the square root of the sample size with which the effect size has been derived as weight.[13] For both the OLS and WLS results Moran's I is not significant, neither for between-study autocorrelation nor for within-study correlation. The LM tests, however, point in the direction of erroneously omitted variables that are correlated within studies. This conforms to expectation in as far considerable effort is often put into specifying differing characteristics of studies, but differences in attributes of estimates of the same study are often neglected, or discarded to save degrees of freedom.

The last two columns of Table 2 reveal the importance of taking into account dependence and weighting to account for the heteroskedasticity inherent in meta-

[13] Preferably one would like to use the estimated standard error of the estimated effect size, but that is unfortunately only available for approximately one-third of the observations. As the variance is inversely proportional to sample size, we use the latter instead.

analysis. Comparison of Tables 1 and 2 makes clear that, as expected, ignoring an autoregressive error structure is much less serious than ignoring the simultaneity with which estimates of the same study have been derived. Uniformly, however, a meta-model that accounts for dependence within studies achieves a better fit than the traditional model ignoring within-study correlation.

7. Conclusions

Meta-analysis is rapidly becoming a well-accepted collection of statistical tools to analyze empirical results of previous studies. It contributes to synthesizing the available knowledge stock, and constitutes a rigorous complement to the traditional state-of-the-art review of the literature. In environmental and natura1 resource economics approximately forty meta-analyses have been conducted over the last two decades.

The validity of meta-analysis crucially depends on finding appropriate ways of dealing with its methodological weaknesses. In environmental economics heterogeneity and to a certain extent selection and publication bias have been dealt with adequately. Between and within-study dependence of estimated effect sizes have, however, largely been ignored, among other things because easy-to-use techniques are lacking. We propose to adapt tools used in disciplines where multidimensional dependencies are frequent, to meta-analysis in economics. The multidimensionality is a pivotal feature of most meta-analyses in (environmental) economics, owing to multiple sampling from the same studies.

In this paper we conceptualize the autocorrelation or dependence problem in meta-analysis, and introduce cross-product statistics as a way of measuring them. We also lay out a framework for testing and visualizing between- and within-study dependence, and incorporate dependence in the meta-regression framework. We illustrate the use of the techniques through a numerical example and two applications.

Several conclusions arise from the above analysis. First, dependence (or autocorrelation) is the rule rather than the exception. Although between-study dependence is usually sufficiently accounted for by means of specifying heterogeneity between studies, within-study dependence typically is not. Second, the implications of ignoring within-study dependence are serious. In particular the consequences of erroneously omitting the simultaneity present among estimates of the same study leads to biased estimators. In general, ignoring dependence will cause inferences about the size and significance of the causes for variation in estimated effect sizes to be inaccurate. Third, hierarchical leve1 models (HLM) constitute an interesting alternative to the models proposed in this paper. Further efforts should be put into modeling dependence in the HLM framework. Finally, this paper provides an example and applications of the proposed framework, but further work in an experimental simulation context (Monte Carlo experiments) is needed to compare the estimators proposed in this paper to estimators allowing for cross-correlation within studies, and estimators based on averaging estimates from the same study.

We emphasize in closing the importance of visualization, and the ease with which it can signal dependence problems. Even the meta-analyst who does not want to engage in extensive misspecification testing should provide Moran scatterplots with trend-lines. A simple correction for studies with only one estimate guarantees that the slope of the trend-line is equal to a correlation-like measure revealing the importance of dependence within and between studies.

References

Anselin, L., *Spatial econometrics: methods and models*, Dordrecht: Kluwer Academic Publishers, 1988a.

Anselin, L., Lagrange Multiplier test diagnostics for spatial dependence and spatial heterogeneity, *Geographical Analysis 20*, 1988b, 1- 17.

Anselin, L., *SpaceStat: a program for the analysis of spatial data*, National Center for Geographic Information and Analysis, Santa Barbara: University of Califomia, 1992 *(see* also, *Space-stat version 1.50: revision notes*, Morgantown: West Virginia University, Regional Research Institute, Research Paper No. 9428, 1994).

Anselin, L., Local indicators of spatial association — LISA, *Geographical Analysis 27*, 1995, 93-115.

Anselin, L., The Moran scatterplot as an ESDA tool to assess local instability in spatial association, in: M. Fischer, H. Scholten and D. Unwin (eds.), *Spatial analytical perspectives on* GIS, London: Taylor and Francis, 1996.

Anselin, L., *Spatial externalities, spatial multipliers and spatial econometrics*, Urbana-Champaign: Regional Econornics Applications Laboratory (REAL) and Dept. of Agricultural and Consumer Economics, University of Illinois, 2001 (mimeo).

Anselin, L. and A.K. Bera, Spatial dependence in linear regression models with an introduction to spatial econometrics, in: A. Ullah and D. Giles (eds.), *Handbook of applied economic stutistics, New* York: Marcel Dekker, 1998.

Anselin, L. and R.J.G.M. Florax, Smal1 sample properties of tests for spatial dependence in regression models: some further results, in: L. Anselin and R.J.G.M. Florax (eds.), *New directions in spatial econometrics,* Berlin: Springer-Verlag, 1995.

Anselin, L., A.K. Bera, R.J.G.M. Florax and M.J. Yoon, Simple diagnostic tests for spatial dependence, *Regional Science and Urban Economics 26,* 1996, 77-104.

Ashenfelter, O., C. Harmon and H. Oosterbeek, A review of estimates of the schooling/ eamings relationship, with tests for publication bias, *Labour Economics 6,* 1999, 453-70.

Boyle, K.J., G.L. Poe and J.C. Bergstrom, What do we know about groundwater values? Preliminary implications from a meta analysis of contingent-valuation studies, *American Journal of Agricultural Economics 76,* 1994, 1055-61.

Brouwer, R., I.H. Langford, I.J. Bateman, T.C. Crowards and R.K. Turner, A meta-analysis of wetland contingent valuation studies, Norwich: Centre for Social and Economic Research on the Global Environment (CSERGE), Working Paper GEC97-20, 1997.

Burridge, P., On the Cliff-Ord test for spatial correlation, *Journal of the Royal Statistical Society B 42,* 1980, 107-8.

Card, D. and A.B. Krueger, Time-series minimum-wage studies: a meta-analysis, *American Economic Review 85,* 1995, 238-43.

Carson, R.T., N.E. Flores, K.M. Martin and J.L. Wright, Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods, *Land Economics 72,* 1996, 80-99.

Cavlovic, T.A., K.H. Baker, R.P. Berrens and K. Gawande, A meta-analysis of environmental Kuznets curve studies, *Agriculture and Resource Economics Review 29, 2000, 32-42.*

Cliff, A.D. and J.K. Ord, Testing for spatial autocorrelation among regression residuals, *Geographical Analysis 4,* 1972, 267-84.

Cliff, A.D. and J.K. Ord, *Spatial processes: models & applications,* London: Pion, 1981.

Cochran, W.G., The combination of estimates from different experiments, *Biometrics* 10, 1954, 101-29.

Cooper, H. and L. Hedges, *Handbook of research synthesis,* New York: Russell Sage Foundation, 1994.

Cressie, *N., Statistics for spatial data, New* York: Wiley, 1991.

Dalhuisen, J.M., R.J.G.M. Florax, H.L.F. de Groot and P. Nijkamp, 2001, *Price and income elasticities of residential water demand: why empirical estimates differ,* Amsterdam: Tinbergen Institute Discussion Paper No. 2001-057/3, 2001.

Espey, M., Explaining the variation in elasticity estimates of gasoline demand in the United States: a meta-analysis, *The Energy Journal* 17, 1996, 49-60.

Espey, M., Gasoline demand revisited: an intemational meta-analysis of elasticities, *Energy Economics 20,* 1998, 273-95.

Espey, M., J. Espey and W.D. Shaw, Price elasticity of residential demand for water: a meta-analysis, *Water Resources Research 33,* 1997, 1369-74.

Espey, M. and D.D. Thilmany, Farm labor demand: a meta-regression analysis of wage elasticities, *Journal of Agricultural and Resource Economics* 25, 2000, 252-66.

Florax, R.J.G.M., Methodological pitfalls in meta-analysis: publication bias, in: R.J.G.M. Florax, P. Nijkamp, K.G. Willis (eds.), *Comparative environmental economic assessment,* Cheltenham: Edward Elgar, 2002a.

Florax, R.J.G.M., *Meta-analysis in environmental and natura1 resource economics: the need for strategy, tools and protocol,* Amsterdam: Dept. of Spatial Economics. Free University, 2002b (manuscript).

Galbraith, R.F., A note on graphical presentation of estimated odds ratios from several clinical trials, *Statistics in Medicine 7,* 1988, 889-94.

Getis, A., Spatial interaction and spatial autocorrelation: a cross-product approach. *Environment and Planning A 23, 199* 1, 1269-77.

Glass, G.V., B. McGaw and M.L. Smith, *Meta-analysis in social research,* Beverly Hills: Sage, 1981.

Greene, W.H., *Econometric analysis,* Englewood Cliffs: Prentice Hall, 1993.

Hedges, L., The promise of replication in labour economics, Labour Economics 4, 1997, 111-14.

Hedges, L. and 1. Olkin, *Statistical methodsfor meta-analysis,* New York: Academic Press, 1985.

Hubert, L., R. Golledge and C. Costanzo, Generalized procedures for evaluating spatial autocorrelation, *Geogruphicul Anulysis* 13, 198 1, 224-33.

Jeppesen, T., J.A. List and H. Folmer, Environmental regulation and new plant location decisions: evidence from a meta-analysis, *Journal of Regional Science,* 2001 (forthcoming).

Katzman, M.T., 1987, Multiattribute utility elicitation techniques and public policy: a meta-analysis of empirical applications, *Applicutions of Management Science 5, 237-303.*

List, J.A. and C.A. Gallet, What experimental protocol influence disparities between actual and hypothetical stated values? Evidence from a meta-analysis, *Environmental und Resource Economics,* 2001 (forthcoming).

Loomis, J.B. and D.S. White, Economic benefits of rare and endangered species: summary and meta-analysis, *Ecologicul Economics* 18, 1996, 197-206.

Moran, P.A.P., The interpretation of statistical maps, *Journal of the Royal Statistical Society, series B* **10,** 1948, 243-5 1.

Moran, P.A.P., Notes on continuous stochastic phenomena, *Biometriku 37,* 1950, 17-23.

Nelson, J.P., Airports and property values: a survey of recent evidence, *Journal of transport economics und policy* **19, 1980, 37-52.**

Rosenthal, R., *Meta-analytical procedures for sociul research,* Beverly-Hills: Sage, 1984.

Schwartz, J., Air pollution and daily mortality: a review and meta-analysis, *Environmentul Research 64,* 1994, 36-52.

Smith, V.K. and J.C. Huang, Hedonic models and air pollution: twenty-five years and counting, *Environmental und Resource Economics 3,* 1993, 381-94.

Smith, V.K. and J.C. Huang, Can markets value air quality?, *Journal of Politica1 Economy* **103,** 1995, 209-27.

Smith, V. and Y. Kaoru, Signals or noise: explaining the variation in recreation benefit estimates, *Americun Journal of Agriculturul Economics 72,* 1990a, 419-33.

Smith, V. and Y. Kaoru, What have we leamed since Hotelling's letter? A meta-analysis, *Economics Letters 32,* 1990b, 267-72.

Smith, V. and L. Osbome, Do contingent valuation estimates pass a "scope" test? A meta-analysis, *Journal of Environmental Economics und Management* **31,** 1996, 287-301.

Stanley, T.D., Wheat from chaff: meta-analysis as quantitative literature review, *Journal of Economic Perspectives* **15, 200 1, 13 1-50.**

Sutton, A.J., K.R. Abrams, D.R. Jones, T.A. Sheldon and Fujian Song, *Methods for meta-analysis in medical research,* Chichester: John Wiley & Sons, 2000.

Woodward, R.T. and Y.-S. Wui, The economic value of wetland services: a meta-analysis, *Ecologicul Economics* 37, 2001, 257-70.

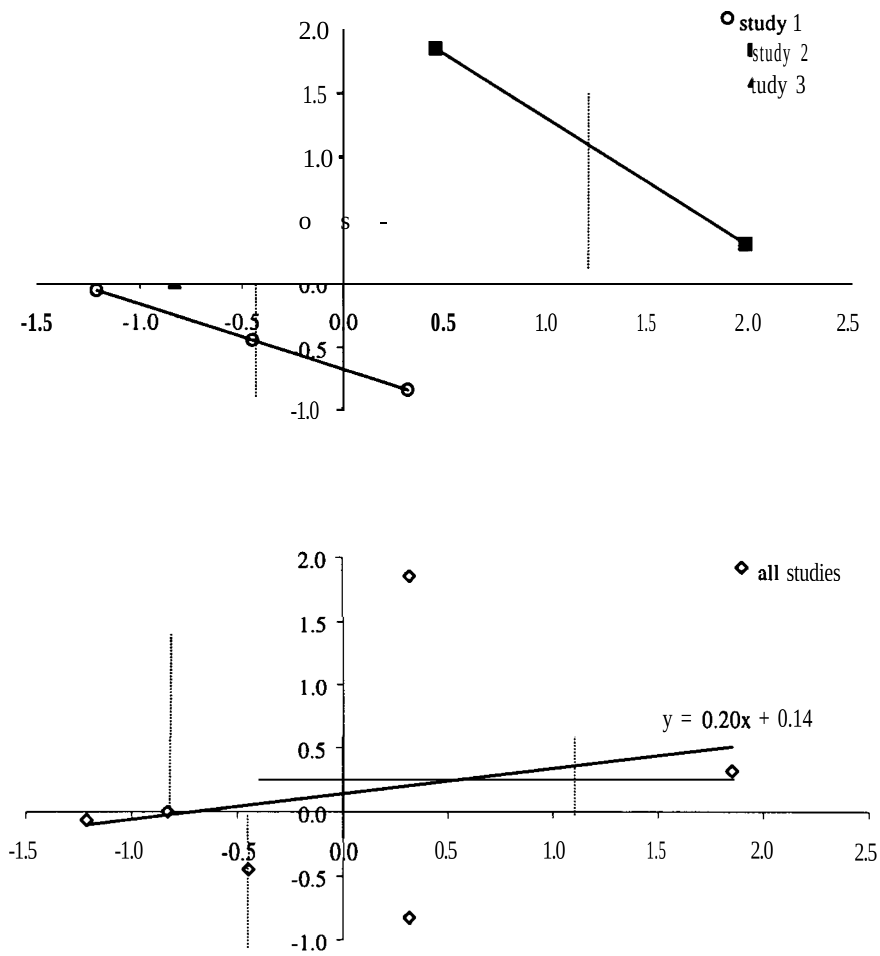

Figure la Scatterplots of within-study dependence, by studies (top) and including all effect sizes (bottorn).
Note Study means are shown as dashed lines.

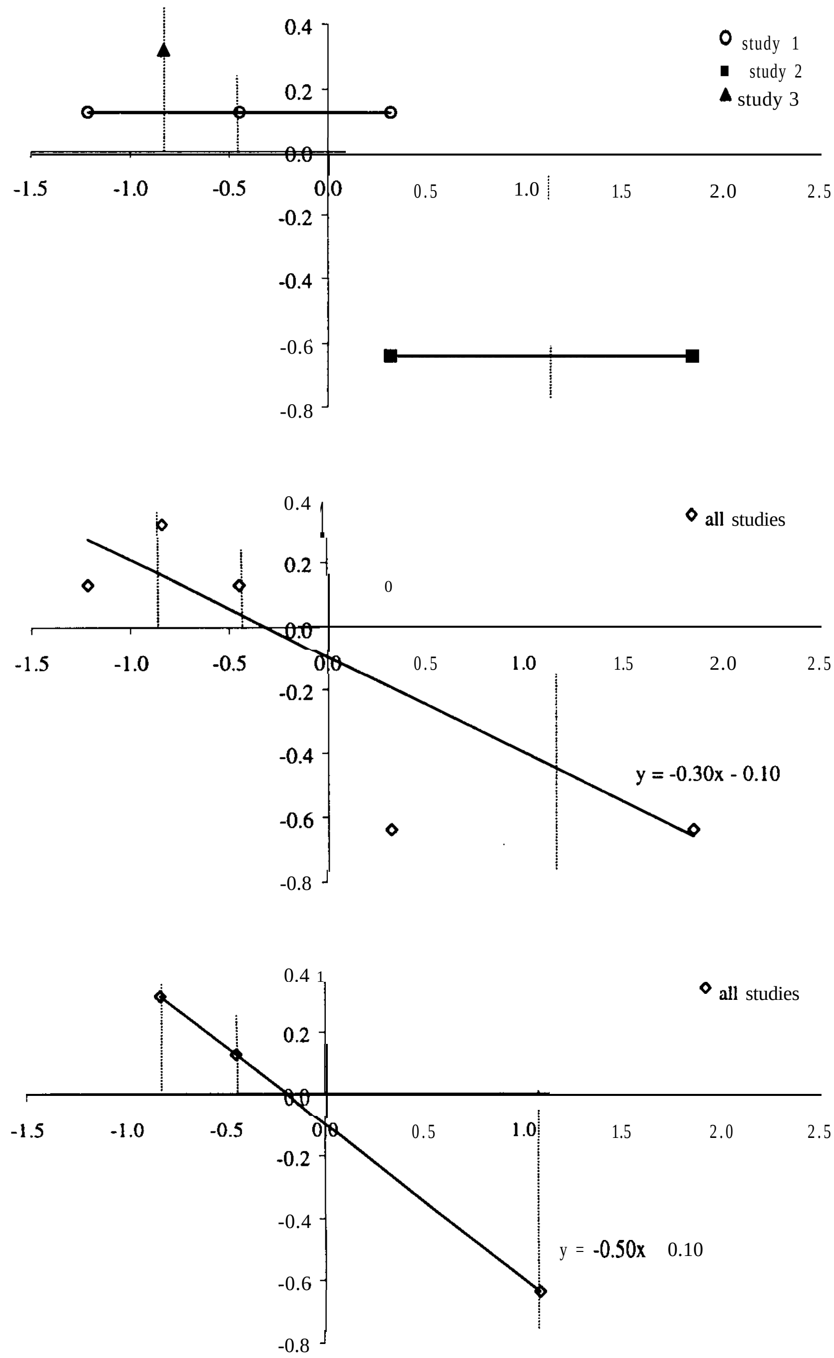

Figure lb Scatterplots of between-study **dependence**, by studies (top), including all effect sizes (middle) and using study means only (bottom).
Note Study means are shown as dashed lines.

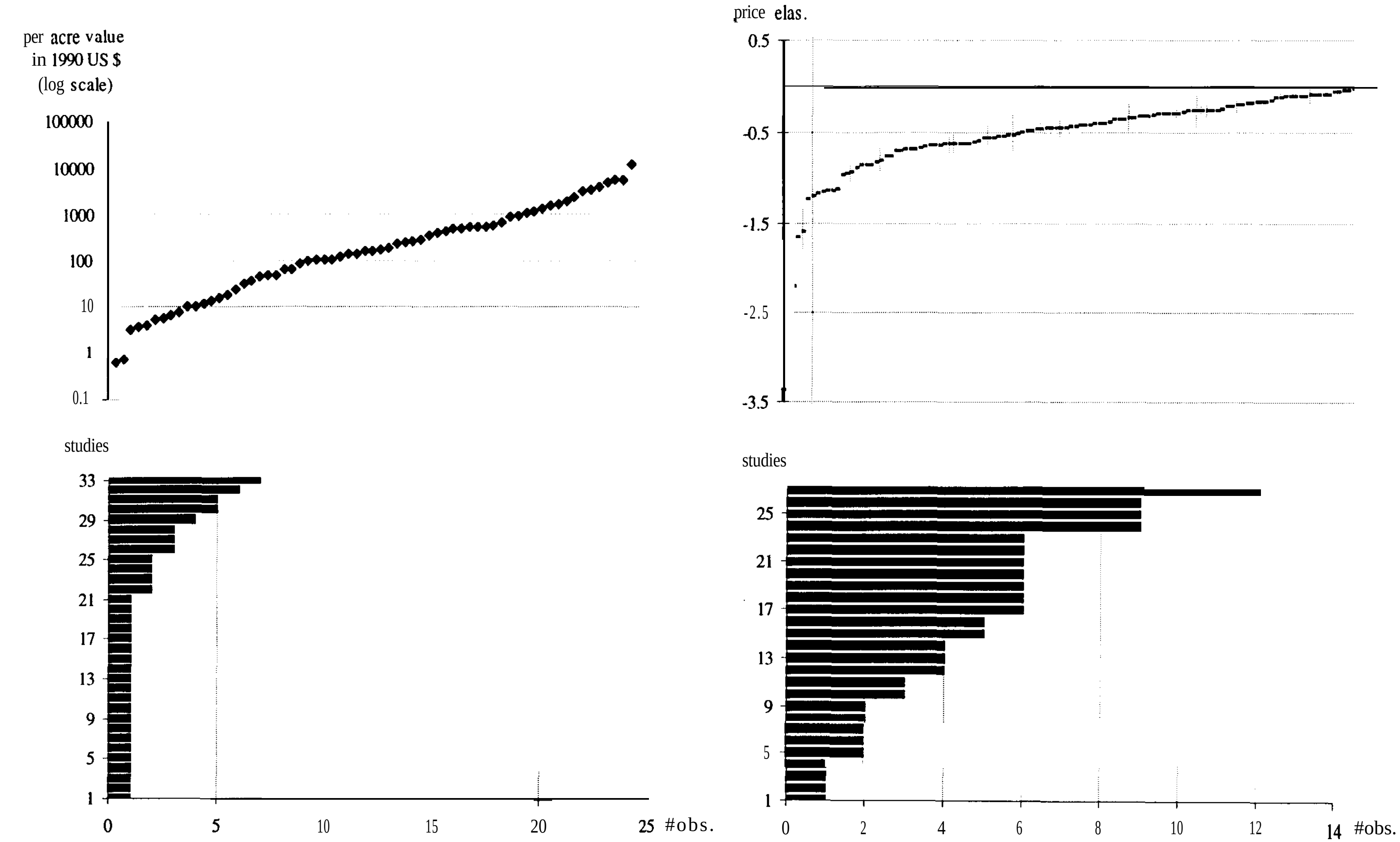

Figure 2 The distribution of effect size estimates, plus and minus one estimated standard error for the price elasticities (when available, top), and the distribution of the number of sampled estimates per study (bottom).

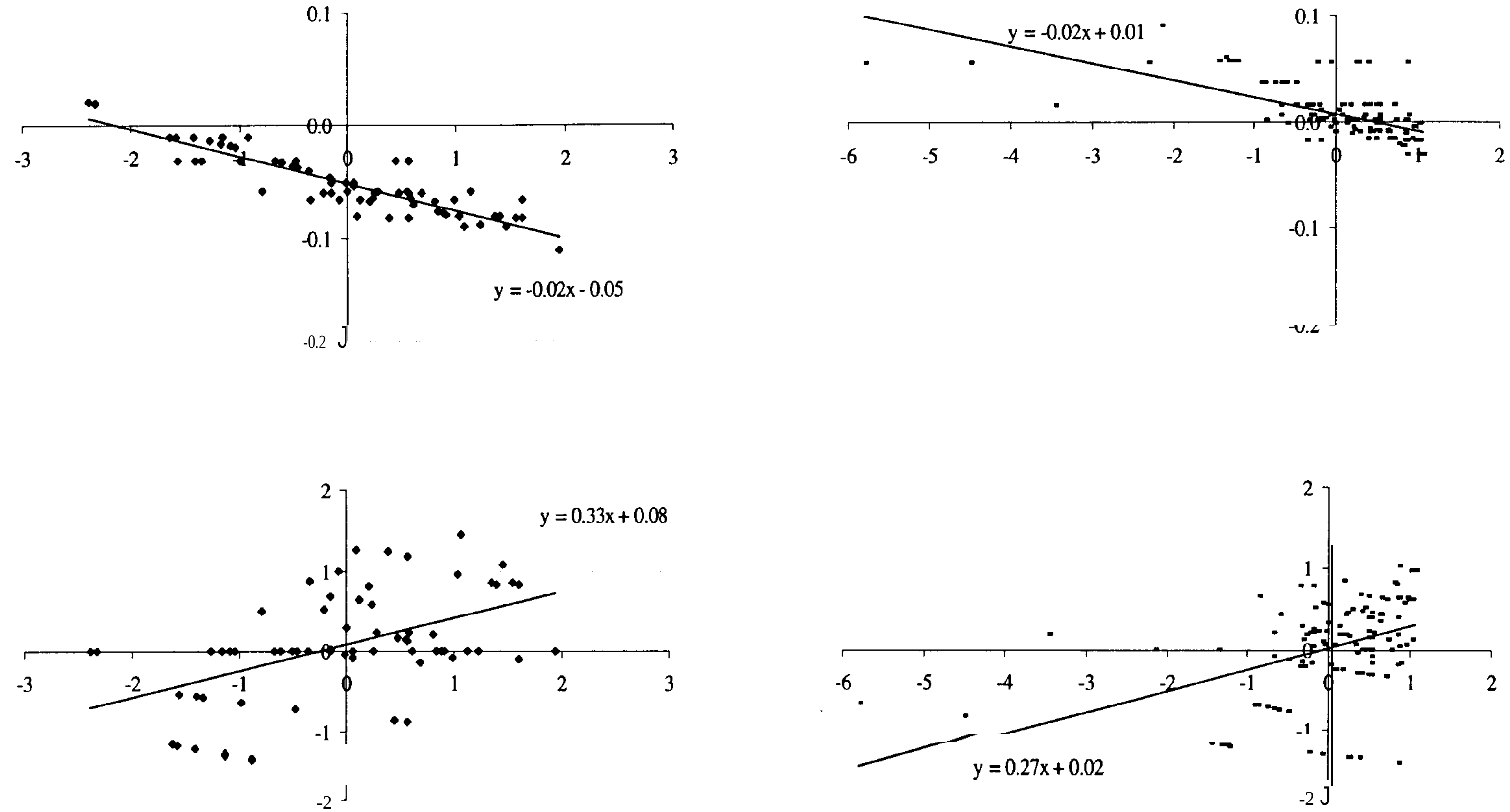

Figure 3 **Moran** scatterplots for within- and between-study dependence (top and bottom, respectively) for the wetlands and water use application (left and right, respectively). *Note* The horizontal axis gives the standardized value of the effect size, and the vertical axis the standardized value of the effect size within or between studies. Standardization implies the usual subtraction of the overall sample mean and division by the maximum likelihood standard deviation of the overal1 sample. The same standardization applies to the original effect sizes (horizontal axis) and the within- and between study effect sizes (vertical axis).

Table 1 Estimated models of wetland valuation functions using OLS and maximum likelihood estimation of a lag model including within-study dependence.[a]

	OLS			*MLLAG*		
				A	*B*	*c*
ρ				0.348***	0.278***	0.303***
				(0.08)	(0.07)	(0.08)
Constant	7.945***	6.641***	7.872***	4.920***	6.004***	5.929***
	(1.38)	(1.44)	(1.87)	(1.30)	(1.19)	(1.43)
Year	-0.052	-0.004	0.016	-0.042	-0.0006	0.006
	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)	(0.03)
Ln Acres	-0.168		-0.286**	-0.065		-0.178*
	(0.11)		(0.12)	(0.09)		(0.09)
Coastal	-0.523		-0.117	0.113		0.532
	(0.88)		(1.02)	(0.70)		(0.75)
Flood	-0.358		0.678	0.329		1.057
	(0.94)		(0.97)	(0.74)		(0.71)
Quality	1.494		0.737	1.627**		0.683
	(0.90)		(1.00)	(0.69)		(0.72)
Quantity	0.514		-0.452	0.860		0.234
	(1.79)		(1.92)	(1.39)		(1.39)
Rec. fish	0.395		0.582	0.172		0.244
	(0.78)		(0.82)	(0.60)		(0.60)
Com. fish	0.669		1.360	1.403**		1.790
	(0.89)		(1.06)	(0.69)		(0.58)
Birdhunt	-1.311*		-1.055	-0.807		-0.701
	(0.73)		(0.80)	(0.57)		(0.58)
Birdwatch	1.704**		1.804**	0.956		1.009*
	(0.75)		(0.80)	(0.59)		(0.59)
Amenity	-3.352***		-4.303***	-3.096***		-3.835***
	(0.95)		(1.18)	(0.74)		(0.86)
Habitat	0.577		0.427	1.245**		0.903
	(0.78)		(0.76)	(0.61)		(0.56)
storm	0.310		0.173	0.994		0.698
	(1.63)		(1.55)	(1.26)		(1.12)
Publish		-0.669	-0.154		-1.175*	-0.394
		(0.77)	(0.85)		(0.64)	(0.62)
Data		0.302	0.0002		-0.391	-0.321
		(0.76)	(0.85)		(0.65)	(0.61)
Theory		-1.020	-1.045		-0.467	-0.769
		(0.82)	(0.91)		(0.70)	(0.66)
Metric		-4.030***	-3.186**		-3.707***	-2.699***
		(1.10)	(1.27)		(0.90)	(0.92)
PS	-2.416**	-2.034**	-3.140***	-2.796***	-1.853***	-3.394***
	(1.02)	(0.85)	(1.03)	(0.79)	(0.69)	(0.74)
HP		0.441	5.043**		-0.768	3.210**
		(1.72)	(2.09)		(1.43)	(1.55)
NFI		-0.724	0.273		-0.579	-0.258
		(1.019)	(1.34)		(0.84)	(0.97)
RC		1.376	2.232**		1.286*	1.943***
		(0.94)	(1.02)		(0.77)	(0.73)
TC		-1.196	-0.341		-1.414*	-0.583
		(0.90)	(1.13)		(0.74)	(0.81)

Table 1 continued.

	OLS			*MLLAG*		
	A	*B*	*C*	*A*	*B*	*c*
R^2 *(adj.)*	0.20	0.25	0.36	0.47	0.45	0.64
F	2.128**	3.084***	2.659***			
Likelihood	-130.001	-130.510	-116.840	-123.008	-124.529	-110.613
CN	17.40	14.74	26.85			
JB	2.409	0.184	1.717			
BP [b]	0.909	0.030	0.463	0.518	0.074	1.676
Between studies						
Moran's I	-0.022	-0.021	-0.020			
LMERR	0.832	0.721	0.685			
LMLAG	0.982	0.795	0.733			
Within studies						
Moran's I	0.328***	0.132*	0.079**			
LMERR	5.001**	0.813	0.289			
LMLAG	12.642***	11.629***	11.504***			

[a] All results are obtained using SpaceStat (see http://www.spacestat.com and Anselin 1992). Standard errors based on OLS and ML are reported in parentheses. Significance is indicated with ***, ** and * for the 1, 5 and 10 percent level, respectively.

[b] Breusch-Pagan heteroskedasticity test for random coefficient variation. We also experimented with other forms of heteroskedasticity.

Table 2 Estimated models of price elasticities of residential water demand using OLS, WLS, and maximum likelihood estimation of an error model including within-study dependence and weighting for heteroskedasticity on the basis of the sample size of the original studies.[a]

	OLS	*OLS, White*	*WLS*	*MLERROR*	*MLERROR, weighted*
Constant	0.395	0.395	0.245	0.515	0.354
	(0.63)	(0.27)	(0.38)	(0.44)	(0.27)
Increasing tariffs	-0.143	-0.143	-0.217*	-0.09 1	-0.185**
	(0.20)	(0.09)	(0.12)	(0.15)	(0.09)
Decreasing tariffs	-0.073	-0.073	-0.125	-0.015	-0.084
	(0.20)	(0.09)	(0.12)	(0.16)	(0.10)
Average/fixed price	-0.193	-0.193***	-0.175*	-0.174	-0.153
	(0.18)	(0.06)	(0.10)	(0.17)	(0.10)
Shin price	-0.127	-0.127	-0.098	-0.121	-0.060
	(0.18)	(0.09)	(0.08)	(0.15)	(0.07)
Conditioned on income	-0.033	-0.033	0.142	-0.133	0.041
	(0.38)	(0.20)	(0.24)	(0.26)	(0.16)
Difference variable incl.	-0.082	-0.082	-0.079	-0.073	-0.059
	(0.17)	(0.08)	(0.08)	(0.12)	(0.06)
Discrete-continuous appr.	-1.042***	-1.042***	-0.968***	-1.114***	-1.021***
	(0.22)	(0.10)	(0.10)	(0.16)	(0.08)
Long-run elasticity	-0.104	-0.104	-0.139**	-0.114	-0.119**
	(0.12)	(0.08)	(0.07)	(0.09)	(0.05)
Segment elasticity	-1.281***	-1.281***	-1.205***	-1.340***	-1.246***
	(0.32)	(0.41)	(0.34)	(0.22)	(0.21)
GDP per capita (x 1,000)	-0.041	-0.041***	-0.033*	-0.047**	-0.038***
	(0.03)	(0.01)	(0.02)	(0.02)	(0.01)
Western US	0.235	0.235***	0.208**	0.265**	0.245***
	(0.15)	(0.07)	(0.09)	(0.10)	(0.06)
Time-trend	0.009	0.009**	0.005	0.010*	0.006
	(0.01)	(0.004)	(0.006)	(0.006)	(0.004)
Cond. household size	-0.187	-0.187**	-0.123	-0.246*	-0.172**
	(0.18)	(0.07)	(0.09)	(0.13)	(0.07)
Cond. seasonal dummy	-0.290	-0.290***	-0.298**	-0.307	-0.297***
	(0.33)	(0.07)	(0.14)	(0.21)	(0.09)
Monthly data	-0.569	-0.569***	-0.473***	-0.637***	-0.547***
	(0.26)	(0.15)	(0.15)	(0.18)	(0.10)
Panel data	0.576**	0.576***	0.437***	0.671***	0.532***
	(0.28)	(0.16)	(0.16)	(0.20)	(0.12)
Summer data	-0.410**	-0.410***	-0.347***	-0.478***	-0.423***
	(0.19)	(0.07)	(0.08)	(0.15)	(0.07)
λ				-0.438***	-0.591***
				(0.13)	(0.11)
R^2 *(adj.)*	0.32		0.43	0.42	0.44
F	4.330***				
Likelihood	-54.784		-15.547	-5 1.680	-8.548
CN	68.36				
JB	1222.679***				
BP	37.114***[b]			5.364**[c]	
Between studies					
Moran's I	-0.001				
LMERR	0.007		0.018		
LMLAG	0.009		0.030		
Within studies					
Moran's I	-0.119				
LMERR	3.096"		7.744***		
LMLAG	1.981		3.851**		

[a] All results are obtained using SpaceStat (see http://www.spacestat.com and Anselin 1992). Standard errors based on OLS, the White adjustment, WLS and ML are reported in parentheses. Significance is indicated with ***, ** and * for the 1, 5 and 10 percent level, respectively.

[b] Koenker-Bassett variant of the Breusch-Pagan heteroskedasticity test for random coefficient variation.

[c] Koenker-Bassett variant of the Breusch-Pagan heteroskedasticity test for the variable 'number of observations.'